stitut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), and Mental Health Research Networking Center (CIBERSAM), Barcelona, Spain; [5]Centre for Psychiatric Research and Education, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

1. Fusar-Poli P, Salazar de Pablo G, Correll C et al. JAMA Psychiatry 2020;77:755-65.
2. Davies C, Cipriani A, Ioannidis JPA et al. World Psychiatry 2018;17:196-209.
3. Bosnjak Kuharic D, Kekin I, Hew J et al. Cochrane Database Syst Rev 2019;11:CD012236.
4. Devoe DJ, Farris MS, Townes P et al. J Clin Psychiatry 2020;81:17r12053.
5. Fusar-Poli P, Davies C, Solmi M et al. Front Psychiatry 2019;11:764.
6. Mei C, van der Gaag M, Nelson B et al. Clin Psychol Rev 2021;86:102005.
7. Pozza A, Dettore D. J Clin Psychol 2020;76:392-405.
8. Fusar-Poli P, Radua J, McKenna PJ et al. Front Psychiatry 2020;11:394.
9. Nelson B, Amminger GP, Bechdolf A et al. Lancet Psychiatry 2020;7:378-80.

DOI:10.1002/wps.20896

Selective outcome reporting and the effectiveness of psychotherapies for depression

Only 40% of trials of psychotherapies for depression published between 2015 and 2018 were prospectively registered, and discrepancies between publications and protocols were noted for 76% of registered trials[1]. It is often assumed that such divergences are the result of intentionally favoring statistically significant findings ("selective reporting"). However, discrepancies could be due to other reasons, such as justified protocol amendments, logistic difficulties or carelessness.

A survey of trials published in high-impact clinical psychology journals over four years[2] identified 27 prospectively registered trials, of which only 13 with a clearly specified primary outcome measure and time of assessment. Among these 13 trials, four contained protocol deviations favoring significant findings (for two others this was impossible to adjudicate). However, it is difficult to reliably estimate the prevalence and impact of selective reporting from investigations of such small cohorts of trials. Therefore, we examined differences in effectiveness associated with selective reporting across a complete cohort of prospectively registered trials of psychotherapies for depression.

We conducted a pre-registered survey (PROSPERO: CRD42019136130) of all randomized trials comparing psychological interventions to control conditions for adult depression which started enrollment after July 1, 2005, when journal registration mandates became widespread[3]. We selected trials from a regularly updated meta-analysis of psychotherapies for depression (https://osf.io/825c6/), using the most recent update (January 1, 2020). We identified matching registrations from the publication, key word searches in public registries, or, failing these, by contact with investigators.

Registration was considered prospective if it occurred within one month of enrollment start. For prospectively registered trials with a pre-specified outcome measure and assessment time point, we examined changes in primary depression outcomes between registries and publications. Potential discrepancies included[4]: a) omission of registered primary outcome (non-reporting); b) addition of new, not registered, primary outcome; c) downgrading of registered primary outcome to secondary; d) upgrading of secondary registered outcome to primary; e) assessment time point changes; f) analysis method changes. Selective reporting was adjudicated for a) or b), and, for other discrepancies, on the basis of the judgement of two independent researchers.

Effect sizes were computed as standardized mean differences (SMDs) between intervention and control for primary depression outcomes at post-treatment or the time point specified as primary, using data from publications. For event data (e.g., response, remission), we computed odds ratios and converted them into SMDs[5]. We pooled effect sizes separately for trials with and without selective reporting, using robust variance estimation with weights from a random effects model, small sample adjustment and an assumed correlation between all pairs of observed effects sizes of 0.8[6]. Analyses were run in Stata/SE 16.1.

We found that, out of 353 randomized controlled trials in the cohort, 185 commenced enrollment after July 2005. Of these, 142 (77%, 95% CI: 70%-83%) were registered. Seventy-five trials (40%, 95% CI: 33%-48%) were registered prospectively, 11 of which (15%, 95% CI: 8%-25%) without specifying outcome measures or assessment time points. Fifty-one trials (68%, 95% CI: 56%-78%) were rated as free from selective reporting. Discrepancies between registries and reports were identified for 19/75 (25%, 95% CI: 16%-37%) trials, of which 13 (17%, 95% CI: 10%-28%) were judged as involving selective reporting. For six trials with an omitted registered primary outcome, we queried primary investigators and received four replies, all explaining that the outcome measure had been dropped out before starting data collection.

The summary effect size was –0.81 (95% CI: –1.25 to –0.38, tau^2= 0.22) for trials with selective reporting, and –0.54 (95% CI: –0.65 to –0.43, tau^2=0.10) for trials without. When analyses were limited to outcomes registered as primary, the effect size in trials with selective reporting was slightly reduced to –0.75 (95% CI: –1.21 to –0.29). Conversely, excluding the six trials that omitted a registered primary outcome led to a considerably reduced effect size for trials with selective reporting (SMD=–0.51, 95% CI: –0.83 to –0.19), closely resembling that of trials without selective reporting. Similarly, excluding the four trials with an added non-registered primary outcome led to a reduced estimate (SMD=–0.62, 95% CI: –1.00 to –0.24) in trials with selective reporting. Finally, analyses restricted to self-report and unblinded measures showed a substantially increased effect size for trials with selective reporting (SMD=–1.02, 95% CI: –1.66 to –0.38), but minimal changes in the effect size for trials without selective reporting (SMD=–0.57, 95% CI: –0.69 to –0.44).

Our findings confirm prior smaller and more circumscribed surveys[1,2], by showing that, even after many journals condition-

ed submission on prior registration, prospective registration is implemented in only 40% of trials of psychotherapies for depression. Among prospectively registered trials, 25% displayed discrepancies between registration and publications, and for 17% we judged these discrepancies as favoring statistical significance. Though relatively few, trials with selective reporting were associated with considerably larger effectiveness, when combined in a meta-analysis. Effect sizes diverged by a SMD of 0.27 between trials with and without selective reporting. For reference, selective publication of trials of psychotherapies for depression has been associated with differences in effectiveness of 0.32[7]. Trials with non-reporting of registered outcomes or addition of non-registered ones emerged as the main drivers of effect size inflation.

These data suggest that lack of prior registration and discrepancies between registration and publications remain common in trials of psychotherapies for depression, and are associated with an inflation of effect sizes in those trials, contributing to the current uncertainties in assessing the outcomes of psychological interventions[8,9].

Clara Miguel[1], Eirini Karyotaki[1], Pim Cuijpers[1], Ioana A. Cristea[2,3]
[1]Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [2]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy; [3]IRCCS Mondino Foundation, Pavia, Italy

Supplementary information on the study is available at https://doi.org/10.17605/OSF.IO/NBYW8.

1. Stoll M, Mancini A, Hubenschmid L et al. J Clin Epidemiol 2020;128:49-56.
2. Bradley HA, Rucklidge JJ, Mulder RT. Acta Psychiatr Scand 2017;135:65-77.
3. De Angelis C, Drazen JM, Frizelle FA et al. N Engl J Med 2004;351:1250-1.
4. Chan A-W, Hróbjartsson A, Haahr MT et al. JAMA 2004;291:2457-65.
5. Chinn S. Stat Med 2000;19:3127-31.
6. Tipton E. Psychol Methods 2015;20:375-93.
7. Driessen E, Hollon SD, Bockting CLH et al. PLoS One 2015;10:e0137864.
8. Cuijpers P. World Psychiatry 2019;18:276-85.
9. Kendall T. World Psychiatry 2019;18:293-5.

DOI:10.1002/wps.20900

Estimating the reproducibility of psychotherapy effects in mood and anxiety disorders: the possible utility of multicenter trials

Estimating the reproducibility of psychotherapy effects is essential. This is particularly crucial for trials with large effects, as the inclusion of false-positive trials can lead to erroneous conclusions about treatment efficacy in research syntheses[1].

Multicenter studies allow researchers to estimate the reproducibility of effects directly across centers under comparable study conditions (e.g., comparable enrollment procedures, inclusion/exclusion criteria, assessment plans). In an important sense, implementation of trials at various centers is close to a direct replication of findings. Accordingly, recent standards recognize the benefit of describing individual center effects in multicenter studies[2].

We aimed to review what we know about center effects in multicenter trials with psychotherapy components for the treatment of mood and anxiety disorders. We examined the extent to which such multicenter trials: a) reported the variability of treatment outcomes for individual centers (i.e., random center effects) and/or b) provided an estimate of the strengths of treatment by center interactions (i.e., fixed center effects)[3].

To obtain a representative sample of recent multicenter studies, we conducted on July 18, 2020 a systematic search of studies indexed between 2010 and 2020 in Medline, PsycINFO and Educational Resources Information Center (ERIC). We used the key words "multicenter or multi-center" combined with "psychotherapy or therapy or counseling" and "depression or anxiety" and publication type "clinical trial" and "adult population". We identified 184 papers, of which 30 referred to treatment outcomes in a multicenter randomized clinical trial (overall 6,638 patients, range 22-1025). Descriptive characteristics of the 30 identified multicenter studies can be obtained from the authors upon request.

In all 30 reports, "multicenter" was mentioned in the title or abstract and in the Methods section. The number of centers ranged from 2 to 30, but in four reports this number was not reported. The majority of the trials investigated treatment efficacy (e.g., changes in symptoms) and four studies investigated economic outcomes (e.g., cost-effectiveness analyses). In 20 studies, at least one significant treatment effect was reported (max. Cohen's d ranged from 0.23 to 3.44).

Only one (3%) out of the 30 studies[4] considered sites a random factor, thereby permitting conclusions about variability in outcomes due to sites in general. Only three (10%) studies[5-7] reported an estimate of the treatment by center interactions. Furthermore, seven studies reported that center effects were "not significant", without further specification of the effect. Among the seven studies with large significant treatment contrasts (max. Cohen's d >0.80), only one[4] reported a statistical estimate of a center effect.

One of the strengths of multicenter studies is the opportunity to estimate the reproducibility of effects. The results of our systematic review indicate that, although studies state clearly that they involve multiple sites and often indicate that this adds to the importance of the trial, they typically do not use the full potential of this design to estimate center effects (either random or fixed), thereby obscuring evidence about reproducibility of effects.

To properly assess the degree to which results are reproducible, we recommend that the authors of multicenter studies report the outcomes for all centers and estimate center effects (i.e., differences in effects amongst centers)[8].

Christoph Flückiger[1], Jessica Paul[1], Peter Hilpert[2], Andreea Vîslă[1], Juan-Martin Gómez Penedo[1], Greta Helene Probst[1], Bruce E. Wampold[3,4]
[1]Department of Psychology, University of Zürich, Zürich, Switzerland; [2]Department of Psychology, University of Surrey, Guildford, UK; [3]Modum Bad Psychiatric Center, Vikersund, Norway; [4]University of Wisconsin, Madison, WI, USA